УДК 681.3

Л.А. Святогор, В.П. Гладун
Институт кибернетики им. В.М. Глушкова НАН Украины, г. Киев
aduis@rambler.ru

Машинное понимание текстов естественного языка: онтологическая парадигма

Выполнен сравнительный анализ двух принципиально разных подходов к семантической обработке текстов и дискурсов: логическая интерпретация и онтологический анализ. В зависимости от семантической сложности текста устанавливается сфера применимости каждого направления. Подробно рассматривается метод онтологического анализа, который опирается на систему знаний и формализованную категорию «смысл». Уточняется понятие «машинное понимание текстов ЕЯ».

Введение

В настоящее время ощущается необходимость определиться с двумя подходами к семантическому анализу естественно-языковых текстов. В первом из них успешно применяются *логические методы* представления и обработки и *интерпретации* текстовых знаний в узких предметных областях (ПрО), которые реализованы во многих поисковых и диалоговых системах. Несколько иной, *логико-лингвистический* подход, который сформулирован в виде модели «***Смысл – Текст***» [1], предполагает выделение в тексте определённых семантических структур, которые, со своей стороны, тесно связаны с морфологическими и семантическими подструктурами. Модель знаний в лингвистической модели явно не задана, она не структурирована и не ограничена рамками ПрО, а результат семантического анализа текста не сформулирован.

В отличие от этих направлений для *содержательного анализа* текстов со сложной семантикой или «гибкой структурой» (например, художественных произведений) требуются принципиально иные методы семантической обработки, которые рассматриваются ниже. Они основаны на использовании контекстного знания. Логическая интерпретация и содержательный анализ текстов естественного языка (ЕЯ) не могут быть противопоставлены друг другу: каждый соотносится со своей областью применения. Если логическая обработка, в принципе, устанавливает непротиворечивость высказываний «логическим знаниям» о предметной области и их выводимость из системы аксиом, то в задаче семантического анализа *слабоформализованных* и разнообразных по тематике текстов ЕЯ на первый план выдвигается *понимание* авторского (коммуникативного) замысла, *извлечение* его из текстового материала и *лексическое оформление*.

С позиций искусственного интеллекта *Д.А. Поспелов* выделил пять уровней машинного понимания текстов различной степени семантической сложности. На первом и втором уровне востребуются только те термины, которые содержатся в тексте непосредственно или могут быть получены из текста путём его логического расширения. На третьем уровне для понимания текста нужно привлечь некоторую дополнительную информацию, которую нельзя вывести логическими методами. Четвёртый и пятый уровни требуют, чтобы к интерпретации текста были привлечены такие понятия, которые в тексте отсутствуют, но накоплены в базах знаний о мире, социуме и используются в предметных областях человеческой деятельности.

В последних двух вариантах толкование текста требует привлечения широкого контекстного материала: знаний о целях коммуникации, о социальной среде, её моральных и культурных ценностях и других системных характеристик.

Цель и содержание данной работы состоят в том, чтобы проанализировать и сравнить между собой два различных подхода к семантической обработке текстов – *логическую интерпретацию* и *контекстный анализ* – и на основании сравнения обосновать онтологический метод семантического анализа, который опирается на триаду «***Текст – Знания – Смысл***». В итоге формализуются категории «*онтологический смысл*» и «*машинное понимание текстов ЕЯ*».

Текст и дискурс

Объектом семантического разбора является текст или дискурс. Необходимо обозначить рамки этих понятий следующими вопросами: Что такое текст? Каково его функциональное назначение? Что мы хотим из него извлечь? Последний вопрос является кардинальным для определения задачи машинного понимания текста.

Для выяснения этих вопросов дадим следующее **определение**. *Текст есть лексическая, грамматически и тематически оформленная, конечная форма передачи замысла автора партнёру по коммуникации – в общей знаковой, языковой и знаниевой среде*.

Современная теоретическая лингвистика опирается на семь критериев текста, сформулированных в 1981 г. Бограндом и Дресслером: 1) связность, 2) целостность, 3) интенциональность, 4) приемлемость, 5) информативность, 6) ситуационность и 7) интертекстуальность [2]. В русле данной работы нас будут интересовать следующие параметры текста: связность, целостность (тематическое единство) и интенциональность (авторский замысел).

В данном исследовании никаких ограничений на тематическое содержание текстов не накладывается. Большинство исследователей считают *функциональным назначением текста* фиксацию и передачу рациональных и культурных знаний в процессе коммуникации людей и сообществ. Текст рассматривают как вербализованную и закодированную средствами языка мысль автора (или его замысел).

На практике сложилось нечёткое разделение текстов на два вида: те, которые функционируют в *узкой* специализированной предметной области (ПрО) или в документообороте, и другие, неформатные тексты с *широким спектром* семантических отношений, например – дискурсы и художественные произведения. Первый вид характеризуется небольшим количеством семантических отношений и достаточно высокой структурируемостью данных. Наоборот, *дискурс* определяется как «*связный текст в контексте многих конституирующих и фоновых факторов – социокультурных, психологических и т.д.*» [3]. Он характерен локальными (микро) и глобальными (макро) семантическими структурами, причём последние реализуются сложными, нелинейными схемами их *смыслового взаимодействия*, которые управляются *авторским замыслом*.

Существо дискурса, его суть определяются не столько количеством включённых в ситуацию объектов, сколько переплетением связей между ними. Связи эти имеют преимущественно ассоциативный характер. Именно в многообразии непредсказуемых отношений между семантически оформленными элементами заключена сложность текста для его восприятия.

Под *сложным текстом* мы будем понимать такой, который нельзя представить конечным числом предикативных отношений в силу существования сюжетных, кореферентных и имплицитных связей между отдалёнными частями (смысловыми блоками) текста. Именно референциональные связи «*отражают коммуникативную установку автора, его интенции, фонд знаний партнёра, коммуникативную организацию высказы-*

вания, отношение к контексту» [3]. Эти связи многообразны и трудноперечислимы, не всегда могут быть установлены логическими рассуждениями, а их совокупность совместно с объектами воссоздаёт фрагмент реальной (или ментальной) действительности.

Отсюда вытекают трудности принципиального характера, когда мы хотим извлечь из текста его главное содержание, то есть – выявить ту коммуникативную цель, которую преследует автор. Предполагая, что эта цель выражена в тексте некоторым «***смыслом***», и что вербализованный смысл отделён от автора, мы тем самым признаём *объективность* данной категории. «*Смысл существует сам по себе как инвариант, который выясняется (устанавливается) во время перевода или пересказа текста*» [3]. Более того, чётко утверждается, что «*сам текст на естественном языке является лишь проводником смысла, а замысел и знания автора лежат во вторичной структуре (смысловой структуре или макроструктуре текста)*»[4].

Самым серьёзным фактором, которому в лингвистике уделено недостаточно внимания, является тот, что текст – как закодированная в языке мысль – «погружён» в систему человеческих знаний о мире, которая опосредовано образует «интеллектуальный фон» и самой мысли, и контекст её изложения.

Логика

Отношение логики к смыслу и смысла к логике

Будем исходить из предпосылки, что в классической логике значение истинности любого сложного высказывания, которое построено из простых (аксиоматических) высказываний и логических связок ***и, или, не***, есть функция истинности простых высказываний, причём аргументы и функции заданы на множестве значений (***истина, ложь***). Центральной проблемой теории является проверка *выводимости* утверждений и их логической *непротиворечивости*. Хотя сами по себе логические высказывания инвариантны к содержательной (смысловой) стороне терминов, предполагается, что от бессмысленных высказываний должны уберегать «правильные» аксиомы, поскольку на стадии их формирования закладывается семантическая связь с внешним миром. Кроме того, подразумевается, что формальные правила вывода достаточно корректны, чтобы не вывести результат за пределы смыслового поля. Всё это, однако, при конечной системе аксиом и при достаточно длинном выводе не уберегает – в момент перехода от логического языка к естественному – от появления «*смысловых парадоксов*».

В логике нередко встречается мнение, что в процессе формального вывода сохраняется «***связь по смыслу***» [5], [6]. Рассмотрим пример. Возьмём две аксиомы, не вызывающие никаких возражений: «***В огороде – бузина***» и «***В Киеве – дядька***». Объединив их импликацией ***если – то***, получим безукоризненно истинное высказывание. Однако порождённая здесь логическая истина демонстрирует как раз факт отсутствия смысла. Разумеется, здесь недостаёт аксиомы, связывающей объекты двух посылок, но мы это обнаружили уже за рамками логики («слишком поздно»). В данном примере недостаточность аксиом вывела нас за пределы смыслового поля. Проблема, однако, заключается не в том, что между *истиной* и *смыслом* нет однозначного соответствия, а в том, что такого соответствия *нельзя требовать*. ***Смысл вообще не является компетенцией логических исчислений***, и в логике это понятие не имеет определения. Смысл, понимаемый как совокупность неформальных связей между частями дискурса, *не измеряется двоичной переменной*. Считать смыслом *имя предиката* некорректно – это имя получает смысловую интерпретацию только в естественно-языковом выражении. В свою очередь, *логические связки* ***и, или, не*** также не имеют никакой другой интерпретации, кроме «математической» – это просто

табличные булевы функции. На пути превращения логики Аристотеля в формальную систему, как отмечает Д.А. Поспелов, *«исчезла логика и победила математика*» [7].

Не всё благополучно обстоит также и с аксиоматикой. Попытка придать семантику даже простому аксиоматическому утверждению обходится дорого: приходится уточнять понятия, затем – уточнять вторичные понятия, и этот процесс лавинообразно нарастает, вовлекая в себя всё новые пласты знания [6]. Именно поэтому «*Логик пока не идёт дальше рассмотрения узкоспециализированных понятийных систем»* [6].

Сделаем уточнение. Несмотря на «размежевание» истины и смысла, мы ни в коей мере не отрицаем их взаимной зависимости в ином – этическом плане. Как отмечают психологи, человек «*выделяет особый смысл, который связан с ориентацией на истину как на бытийную ценность и является нормативным по отношению к профессиональной деятельности учёного*» [8].

Сформулируем кратко итог. ***Истинность есть мера выводимости высказываний, а смысл есть мера знания о мире.*** С позиций искусственного интеллекта можно считать, что категория смысла универсальнее истины.

В следующем разделе рассмотрен пример логико-предикативной модели представления знаний, используемой для машинного диалога. Он необходим для сопоставления с онтологической моделью.

Принципы логической обработки знаний

Типичные черты предикатно-логического подхода присутствуют в поисковой системе, подробно и ясно описанной в работе [5]. Предметной областью служит организация производства, снабжения и сбыта в металлургической отрасли Украины (фрагмент отрасли).

Информация о ПрО представлена в *базе знаний* в виде семантических структур ***фактов***, причём данные структуры явно определяют стандартные связи между объектами. Фактом называют *локализованную минимальную совокупность понятий естественного языка, отражающих связь между классами объектов, классами и их элементами, а также между элементами различных классов.* Факты называются также суждениями или высказываниями. Классы объектов перечислены в *базе данных*. Семантическая структура факта представлена деревом, в узлах которого находятся понятия, а дуги отражают семантические отношения между ними. Одновременно такая графическая структура транслируется в предикатную форму.

Запрос, который формулируется на естественном языке с незначительными ограничениями на синтаксис, преобразуется через предикатную форму в семантическую структуру, подобную структурам базы знаний. После этого производится ***распознавание*** структуры запроса в базе знаний. При совпадении структура базы знаний преобразуется в предикат, а последний транслируется в ответ на естественном языке.

Поисковая система поддерживает составные и сложные вопросы. Она может не только продублировать вопрос с выдачей искомых данных, но и представить прогнозы относительно ресурсов, выполнения плана и другую ценную информацию, не представленную в явном виде.

Как утверждает автор рассматриваемой работы, предикат типа: **«Находится (предприятие (A), город (B))»** «*связывает понятия «предприятие» и «город» в **единое, имеющее смысл целое***» [5]. С этим можно согласиться, если рассматривать данную ситуацию как «*функциональное отношение между предикатом и актантами, которое отражает **семантическую структуру** данной ситуации*» [3]. И при этом добавить, что *смысл «целого»* проверяется за пределами логики.

В описанной системе логика предикатов, дополненная структурами знаний, служит естественным аппаратом для проверки истинности выводимых утверждений:

предметная область в понятийном плане ограничена, данные структурированы, предикатные отношения перечислимы (сведены в таблицу), синтаксис упрощён. Однако, как справедливо отмечает Д.А. Поспелов, «*возникают огромные технические трудности, которые перерастают в принципиальные препятствия, как только модель мира становится достаточно объёмной, а число тем, по которым человек хочет вести диалог с ЭВМ – велико*» [7].

В новейших работах логико-семантического направления наблюдается стремление больше *сблизить логику с семантикой*. Для этого, например, в работе [9] в теорию вводятся «*модели миров*». Успех в данном направлении, по нашему мнению, может быть достигнут, если номинативные структуры логики будут связаны с системой репрезентации знаний, а выводимость будет означать обнаружение формальных смысловых связей. Тогда семантическая нагрузка может быть возложена в моделях миров на интенсиональные и экстенсиональные связи, а прерогативой логики останется вывод и синтез сложных семантических понятий из простых.

Что мы хотим извлечь из текста?

В конечном итоге, все методы семантического анализа направлены на формализацию и обработку знаний. Однако при решении задач семантической обработки текстов вырисовывается принципиальная разница не столько в методах, сколько в целевой установке обработки знаний: если логический аппарат ориентирован на вывод и получение *новых* знаний, то семантический анализ текстов нацелен на извлечение *готовых знаний* и их перекрёстный анализ. При этом, как указывалось, *истинность или ложность обнаруженных в тексте конкретных знаний не являются целью семантической обработки*: приоритетная цель заключается в установлении *смысловых отношений* между лексическими элементами, кодирующими знание.

В зависимости от степени формализации текста и сложности мы можем извлечь из него либо ту информацию, которая в нём содержится – в терминах непосредственно составляющих, либо другую информацию, которая раскрывает содержимое терминами контекстного знания. В последнем случае необходимо построить систему контекстного знания – онтологию. Как будет показано ниже, целью семантической обработки является извлечение из текста и дискурса онтологического смысла. Тем самым категория «смысл», извлечённая из ментальной сферы, приобретает определённость и конструктивность, что позволяет вести обработку машинными средствами. Однако, как справедливо указывает И.Б. Штерн, «*рассматривая понимание естественного языка как оперирование смыслом, прежде всего надо разрешить принципиальные проблемы, связанные с* ***определением сущности и структуры смысла***» [3]. Один из возможных способов решения этих проблем предложен в следующих разделах.

Краткие выводы. В тех текстах, где данные могут быть структурированы и ограничены по множеству объектов и отношений (например, в рамках ПрО), целью диалога или поиска служит извлечение данных. В сложных текстах требуется извлечь смысл как цель коммуникации.

Онтология и смысл

Онтологическая картина мира

Наша задача заключается в том, чтобы ***перевести понятие «смысл» из ментальной или подразумеваемой категории – в конструктивную.*** Это возможно сделать лишь в том случае, если смысл будет включён в формальную систему репрезентации знаний – семантическую сеть; наиболее пригодной для наших целей является ***онтология*** [10].

Известны онтологии двух типов – абстрактные и предметные онтологии. *Абстрактные онтологии* (Дж. Совы, СУМО, Микрокосмос и др.) создаются с философских позиций, чтобы ответить на вопрос: какие категории познания создаёт человек для понимания универсума – природы, космоса, материи. Однако на уровнях межличностного и социального общения человек, стремясь быть понятым, предпочитает структурировать мир иначе. В этом случае для передачи коммуникативных знаний подходят *предметные онтологии*, которые лучше передают ситуации и события в окружающей среде. Предметная онтология способна сочетать в себе, в терминологии Ю.Д. Апресяна, «научную» и «наивную» картины мира.

Предметная лексическая онтология представляет собой семантическую сеть, в которой узлам соответствуют языковые понятия, описывающие объекты или состояния – *концепты*, а связи отображают различные *отношения* между ними. В соответствии с законами познания внешнего мира сеть имеет *иерерархическую* архитектуру, где чем больше уровень общности (или объём) понятия, тем более высокое место в структуре оно занимает. «*На стадии концептуализации эта иерархическая структура превращается в* ***пирамиду знаний***» [4]. Формальные концепты и отношения сети имеют под собой мощную *систему интерпретации* на естественном языке.

Для репрезентации глобальной системы всеобщих знаний, ориентированной на анализ текстов с произвольной тематикой, разработана предметная онтология **ИО*3;** она подробно и вместе с примерами описана в работе [11].

Трёхуровневая иерархическая онтология ИО*3

Основными принципами построения трёхуровневой иерархической онтологии **ИО*3** являются:

– структурирование мира на основе материалистической парадигмы академика *В.И. Вернадского*;

– введение *трёх уровней* иерархии («слоёв знаний»), внутри которых развиваются и структурируются многочисленные подуровни концептов;

– выбор в качестве *концептов* имён, которые именуют (кодируют) понятия, объекты, ситуации, состояния и атрибуты в языковой среде общего знания и общения;

– выбор в качестве концептуальных *связей*: формальных, атрибутивных и ассоциативных отношений;

– задание *интерпретирующей системы* на множестве декларативных знаний, хранящихся в учебниках, энциклопедиях и толковых словарях;

– замыкание концептов на *словарь* естественного языка.

Верхний уровень иерархии образует дерево, вершиной которого служит глобальный концепт *Материя*. Далее он раскрывается через понятия: *Косное вещество, Живое вещество*; последнее разделяется на *Биосферу* и *Ноосферу,* и далее деление продолжается, переходя к концептам среднего уровня. Здесь находятся понятия, которые сложились в науке для изучения природы, человека и общества по отраслям знаний. Наконец, общенаучные понятия раскрываются на нижнем уровне через конкретные характеристики и отношения бытового уровня, которые могут быть спрогнозированы в корпусе текстов. Кроме того, на третьем уровне формулируются и подключаются к среднему уровню концепты предметных областей, которые могут, в принципе, представлять любые области знаний и деятельности человека.

Построение пирамиды онтологии подчинено чёткому принципу репрезентации знаний через *интенсионалы* и *экстенсионалы.* Каждое понятие в семантике рассматривается в единстве этих двух категорий. Это значит, что каждый интенсионал должен быть раскрыт соответствующим ему по семантическому отношению экстенсионалом. Практически это означает, что каждый концепт, выступающий в роли

интенсионала или экстенсионала, имеет *значение* из множества ***имён*** объектов или ситуаций, которые имеют ***денотат*** в реальном мире и лексический ***смысл*** в естественном языке [12]. В онтологии **ИО*3** в роли экстенсионала могут выступать: *части понятия* (при классификации), его *признаки* (при атрибуции), *значения* (при уточнении понятия), *примеры* (в качестве представителей класса), а также *ассоциативные понятия* (по усмотрению эксперта). Очевидно, что при движении по вертикали пирамиды (вниз) экстенсиональные компоненты становятся интенсиональными по отношению к собственным определителям. Экстенсионалы через свои имена (значения концептов) взаимодействуют со словарём естественного языка.

Для того чтобы придать пирамиде знаний конструктивные, процедурные свойства, воспользуемся понятием «графа концептуальных зависимостей», введенным ***Р. Шенком***. Уникальным свойством данной конструкции является тот факт, что «***граф концептуальных зависимостей служит метаязыком для внутренней смысловой репрезентации текстов, который отображает смысловую структуру ситуации***» [3].

Иерархическую трёхуровневую онтологию – **ИО*3** будем считать ориентированным *концептуальным графом*, все пути на котором ведут из любого концепта к глобальной вершине *Материя*.

Относительно количества концептов и внутренних связей онтологии никаких ограничений не делается: её структура является *расширяемой* за счет добавления новых знаний, в том числе разных ПрО.

Концептуальный онтологический граф **ИО*3** имеет следующие преимущества, которые выделяют его среди других систем репрезентации знаний:

– граф служит инструментом *формализации* понятия «смысл текста естественного языка»;

– он позволяет свести процедуру семантического анализа текста к *извлечению смысла* и реализует данную процедуру *стандартными графическими операциями*;

– допускает *машинное представление* и машинную обработку текстовых знаний.

Семантические процедуры извлечения из текста его глубинного содержания будем в дальнейшем называть *онтологическим анализом*. Результатом онтологического анализа служит *онтологический смысл*. В следующем разделе показано, каким образом онтологический анализ (выделение смысла) может быть реализован алгоритмически.

Онтологический смысл: формальное представление

В искусственном интеллекте задача извлечения знаний из текстов конкретизуется как «***понимание и выделение смысла текста***» [4]. Как было показано выше, фундаментальной основой для выделения смысла может служить онтология. Однако без формального определения смысл остаётся объектом ментального рассмотрения в психологии и литературоведении. Ранее отмечалось [1], что попытка построить лингвистическую модель «***Смысл – Текст***» не привела к конструктивным результатам, поскольку, используя формальные семантические, предикатные, синтаксические и морфологические структуры, она в то же время не опиралась на систему знаний. Поэтому нашей первой задачей было дать формальное определение смысла как категории мышления и сделать смысл ***вычислимым***.

Полное рассмотрение дано в работе [10]; здесь приводятся только основные определения.

Определение 1. *Конструктом смысла* является пара концептов, связанных между собой формальным, атрибутивным или ассоциативным отношением.

Определение 2. Онтологический смысл строится из конструктов смысла.

Онтологическим смыслом некоторого концепта является подграф концептуального графа онтологии ИО*3, определяемый начальным* и *финальным концептами. Указанный подграф называется ***смысловой траекторией*** начального концепта.

Определение 3. Значением начального концепта смысловой траектории служит *ключевое слово* текста; значением финального – вершина онтологии *Материя*.

Определение 4. Совокупность смысловых траекторий всех ключевых слов текста формально определяет *онтологический смысл текста*.

Смысловой анализ ЕЯ текста начинается с извлечения из него очередного ключевого слова. Основанием для такой операции служит определение, предложенное в работе [4]: «*Набор ключевых слов – это набор опорных точек, по которым развёртывается текст при кодировании в память и осознаётся при декодировании, это семантическое ядро цельности*». Со своей стороны, мы опираемся на гипотезу, что ключевые («значимые») слова являются основным семантическим инструментом автора текста для кодирования своего замысла.

Таким образом, ***формальный онтологический смысл*** получается в результате проекции дискретизированного текста на онтологию **ИО*3**. Результатом отображения в сети каждого ключевого слова служит ***смысловая траектория*** – *дискретная упорядоченная цепочка взаимосвязанных концептов,* которая «считана» с подграфа онтологического графа. Задание онтологического смысла в таком формате пригодно для компьютерного представления и «машинного понимания».

Онтологический смысл отличается от других известных определений, а тем более – от интуитивного понимания смысла, тем, что он *задан в многомерной лексической координатной системе представления знаний о внешнем мире и имеет – в фиксированной онтологии – однозначное графическое представление.*

Машинное понимание текстов

Понимание и связность

Как отмечалось выше, многие исследователи в области искусственного интеллекта определяют ***понимание реалии*** как «*постижение её смысла через конституирующие её контексты*» [13]. *Рассматривая понимание естественного языка как оперирование смыслом, прежде всего необходимо разрешить принципиальные проблемы, связанные с определением сущности и структуры смысла*» [3].

Такие принципиальные проблемы (как было показано выше) имеются. На пути их разрешения нами предложено понятие «онтологического смысла»: именно благодаря ему сущность и структура данной гносеологической категории приобретают формальный вид. Подчеркнём ещё раз, что онтологический смысл не есть отображение высказывания на множество *ноль* и *единица*. Это есть *отображение высказывания на связное множество именованных концептов, репрезентующее систему знаний* – онтологию.

В самом общем случае ***понимание*** является целью коммуникации и прерогативой двух интеллектов. На ментальном уровне процесс понимания означает возбуждение в памяти человека моделей реальности (языковых и образных), нахождение аналогий или закрепление новых связей; на машинном уровне факт «компьютерного понимания» означает возбуждение на графе онтологии соответствующего подграфа смысла и связывание его с другими подграфами. В том и другом случае реконструируется некоторая языковая ситуация (реалия).

Соединение онтологии со смыслом позволяет *соединить семантику и прагматику текста в единый процедурный комплекс*. Действительно, онтология знаний отвечает за связь текста с внеязыковой действительностью, а смысловая траектория отвечает за кодирование онтологического смысла в машинной памяти. Такой синтез семантики и прагматики позволяет нам дать следующее определение: ***машинным пониманием текста называется процесс построения семантических траекто-***

рий, возбуждаемых текстом на графе онтологии, которая во внутренней памяти машины выполняет роль системы репрезентации знаний о внешнем мире.

Таким мы видим основной результат онтологического анализа ЕЯ тестов.

Ещё один полезный результат от введения онтологического смысла состоит в уточнении термина ***связность текста***. Связный текст есть функция связной мысли. Однако, будучи отделён от мысли, текст, особенно дискурс, имеет дискретную тематическую или сюжетную структуру. В литературоведении под связностью (например, повести) понимают возможность наблюдать за развитием выделенного сюжетного фрагмента в контексте полного сюжета, отслеживая определённую тему на фоне других. Связность дискурса реализуется авторскими приёмами, чаще всего – через имена действующих лиц, место и время действия. На уровне формальной логики связность текста полностью достигается переносом слова из предыдущего предложения в последующее, т.е. связность умозаключений является, в определённом смысле, синонимом выводимости.

Благодаря формализации смысла в системе знаний можно ввести конструктивное понятие ***связность текста по онтологическому смыслу***. Два ключевых слова в тексте порождают на графе две смысловые траектории, которые либо *сходятся* в одну точку – на вершине пирамиды, либо *пересекаются* на одном из трёх уровней онтологии **ИО*3**. Первый общий концепт этих траекторий говорит о том, ***что*** связывает два ключевых слова на концептуальном уровне. Именно этот общий концепт (общее понятие в языке) *именует ту ситуацию*, в которой ключи вступают в *первое взаимодействие*. Чем выше в пирамиде знаний происходит первое замыкание траекторий, чем выше уровень абстрагирования пары слов из текста, тем менее связным является изложение.

Краткие выводы. На уровне графических структур мы соединили семантику с прагматикой, «заставив» ***машину понимать смысл.*** «Понимание» и «смысл» соотносятся как синонимы. Связность текста устанавливается через пересечение траекторий смысла.

Зачем нам нужен онтологический смысл?

Не отрицая ментальное происхождение смысла, которое подтверждается утверждением, что «*смысловым содержанием фраза наполняется в психическом процессе*» [12], мы можем дополнить: «***смысловое содержание фразы выявляется в онтологическом процессе машинного понимания теста***».

Какие последствия может иметь введение «онтологического смысла» для теории и практики?

В теоретическом плане – открыта научная методология глубинного анализа естественно-языковых текстов сложной семантики через синтез онтологического смысла. ***Смысл становится вычисляемой семантической функцией текста и дискурса.*** В поле знаний об искусственном интеллекте разработана цельная концепция: «текст – знания – смысл», а в инженерии знаний сделан шаг по «***превращению искусства в ремесло***».

В практическом плане – извлечение из текста онтологического смысла позволяет разгрузить пользователя сети Интернет от рутинной работы по анализу потоков ненужной (нерелевантной запросу) информации. «Смысловой фильтр» онтологии не только отсеет «информационный мусор», но и представит текст в сжатом виде, сохранив только его смысл. Благодаря этому в системах поддержки принятия решений руководитель или менеджер получает возможность быстро вникнуть в суть проблемы – ***не читая сообщение полностью, а просмотрев его «смысловой портрет».***

В перспективе пользователь сможет построить собственную онтологию сферы его интересов и в рамках ***«персональной онтологии» знаний*** оперировать актуальной информацией на смысловом уровне.

Литература

1. Мельчук И.А. Опыт теории лингвистических моделей «Смысл – Текст» / Мельчук И.А. – М. : Школа «Языки русской культуры», 1999. – 346 с.
2. Методы анализа текста и дискурса / С. Тичер, М. Мейер, Р. Водак, Е. Веттер. ; [пер. с англ.]. – Х. : Изд-во Гуманитарный Центр, 2009. – 356 с.
3. Штерн І.Б. Вибрані топіки та лексикон сучасної лінгвістики. Енциклопедичний словник / Штерн І.Б. – К. : «АртЕк», 1998. – 336 с.
4. Гаврилова Т.А. Извлечение и структурирование знаний для экспертных систем / Т.А. Гаврилова, К.Р. Червинская. – М. : Радио и связь, 1992. – 200 с.
5. Одинцов Б.Е. Пресуппозиция в ассоциативных сетях и машинное понимание текстов / Одинцов Б.Е. – Львов : Львовский госуниверситет им. И. Франко. – 1996. – 60 с.
6. Рубашкин В.Ш. Представление и анализ смысла в интеллектуальных информационных системах / Рубашкин В.Ш. – М. : Наука, 1989. – 192 с.
7. Поспелов Д.А. Фантазия или наука: на пути к искусственному интеллекту / Поспелов Д.А.– М. : Наука, 1982. – 224 с.
8. Балл Г.А. Психология в рациогуманистической перспективе : избранные работы / Балл Г.А. – К. : Основа, 2006. – 408 с.
9. Нікітченко М.С. Математична логіка та теорія алгоритмів : підручник / М.С. Нікітченко, С.С. Шкільняк. – К. : Видавничо-поліграфічний центр «Київський університет», 2008. – 528 с.
10. Святогор Л. Определение понятия «Смысл» через онтологию. Семантический анализ текстов естественного языка / Л. Святогор, В. Гладун. // International Book Series, Number 9. Intelligent Processing. Supplement to the International Journal "Information Technologies & Knowledge". – 2009. – № 3. – P. 53.
11. Гладун В. Структурирование онтологии ассоциаций для конспектирования естественно-языковых текстов / В. Гладун, В. Величко, Л. Святогор. // International Book Series, Number 2. Advanced Research in Artificial Intelligence. Supplement to the International Journal "Information Technologies & Knowledge". – 2008. – № 2. – P. 153.
12. Шевченко А.И. Актуальные проблемы теории искусственного интеллекта / Шевченко А.И. – Донецк : «Наука і освіта», 2003.
13. Штерн И.Б. Интродуктивные модели гуманитарных знаний: концептуальне гештальты versus понятия / И.Б. Штерн // KDS–97. Шестая Международная конференция «Знания – Диалог – Решение» : сборник научных трудов в двух томах. – Ялта, 1997. – Т. 1. – С. 89.

Л.О. Святогор, В.П. Гладун

Машинне розуміння текстів природної мови: онтологічна парадигма

Зроблено порівняльний аналіз двох принципово різних підходів до семантичної обробки текстів та дискурсів: логічна інтерпретація і онтологічний аналіз. Залежно від семантичної складності тексту встановлюється сфера вживання кожного напрямку. Детально розглядається метод онтологічного аналізу, що спирається на систему знань та формалізовану категорію «смисл». Уточнюється поняття «машинне розуміння текстів природної мови».

L.A. Sviatogor, V.P. Gladun

Machine Understanding of A Natural Language Texts: an Ontological Paradigm

The comparative analysis of two essentially different approaches to semantic processing texts and discources is executed: logical interpretation and ontological analysis. Depending on semantic complexity of text the sphere of applicability of each direction is established. The ontological analysis method which leans on knowledge representation system and the new formalized category "the ontological sens" is in detail considered. The concept "machine understanding" of natural language texts is constructively determined.

Статья поступила в редакцию 02.07.2010.